PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Validation of intellectual disability coding through hospital morbidity records using an intellectual disability population-based database in Western Australia
AUTHORS	Bourke, Jenny; Wong, Kingsley; Leonard, Helen

VERSION 1 – REVIEW

REVIEWER	Professor Roy McConkey Ulster University, Northern Ireland, UK
REVIEW RETURNED	22-Aug-2017

GENERAL COMMENTS	This paper is predicated on a curious assumption that hospital morbidity data would provide a good indicator of the prevalence of ID in a childhood population (Page 4 lines 42 to 49). The authors cite no past research to support this proposition although the studies they do reference, indicate that only a minority of children known to have ID will have been admitted to hospital (albeit at a higher rate than their non-ID peers). Indeed their own study bears this out as shown in Figure 1; with around 18% of children with ID having a recorded hospital admission in a 27 year period. Moreover the small number of children aged over 12 months identified solely in hospital admissions (n=354) would make little difference, I suspect, to the prevalence rates of ID within Western Australia. This is information that the authors do not include although it gets a fleeting mention in the discussion. Thus the extensive (and presuming time-consuming) data analysis only serves to confirm an unlikely proposition (acknowledged by the authors on page 8; lines 55 – 57 and again on page 12, lines 43-48). The detailed comparisons that the authors present of the 'cause of disability' codings within the hospital sample and the larger ID population (Table 2) is of limited usefulness as no information is provided as to who undertook the codings in the two databases and the criteria used. Also it seems that only one 'cause code' was entered per case (based on totals given in Table 2) which can be a further source of inconsistency across the two systems. Indeed the authors go on to recognise in the discussion two further sources of error in the hospital codings: some children with ID will have been missed and children with particular syndromes may be assumed to have ID when this may not have been the case. Likewise the IDEA system misses children with ID who do not use/require services; especially younger children undergoing assessment.

	In sum, this study seems misdirected and not a good use of the human and financial resources incurred in data linkage studies. Rather a better focus would have been to explore the number and nature of hospital admissions of children with ID and possible changes over the years. Also the implications for primary care services that might have prevented or reduced the need for admissions perhaps could be identified as well as the implications for improved care in hospital.

REVIEWER	Sally-Ann Cooper University of Glasgow, UK
REVIEW RETURNED	09-Sep-2017

GENERAL COMMENTS	The study is clearly written and interesting. There are two main issues that the authors should address prior to publication. 1. It is obvious that hospital admission data cannot provide a credible population prevalence for intellectual disabilities. There are several reasons for this, including, a. only some infants/children require hospital admission, b. those that do are likely to have the reason for their admission recorded rather than whether they have intellectual disabilities (which one would expect to be recorded more hapazardly if at all), and c. the infant's/child's intellectual disabilities may not have been identified by the time of admission as some are not identified until school age. So what the study achieves is a quantification of the extent to which hospital admission is an inadequate source for identification of population prevalence of intellectual disabilities - they found only 14% of children thought to have intellectual disabilities could be identified in this way. I think the study would be more credible if the authors are clearer about this. In particular, paragraph 2 of the introduction needs rewriting. The first half of the sentence "on account of high likelihood of hospitalisations health data have often been used as a source of ascertainment for prevalence of intellectual disabilities, as have other administrative datasets relating to education or service provision as well as household sampling.." has not been justified by the references cited, and I suspect is actually incorrect. That is, the first part of this sentence on health data has not been justified/is not correct (contrary to the statement regarding administrative education and other service datasets and household sampling which is quite different and is justified by the meta-analysis quoted, and indeed the "gold standard" used in this study (IDEA) relates to administrative education and service data). The authors should rewrite this paragraph for greater accuracy so that their study has more credibility overall. 2. As the authors comment on in the discussion, the codes they use in HMDS for identification of children who might have intellectual disabilities are over-inclusive as not all children with these conditions will have intellectual disabilities e.g. marfan's syndrome. They should therefore also be more precise in their descriptions/reporting in tables 2 and 3 and the text relating to these tables. Table 2 "children in the IDEA database……..identified/not identified with intellectual disabilities in HMDS….." does not provide data on those identified/not identified with intellectual disabilities in HMDS, it provides data on those with the codes they have used as potential proxies for intellectual disabilities - unless I have misunderstood the methods/the methods have not been clearly written.

	Please clarify if these children have been identified/not identified with intellectual disabilities or if this refers to the longer list of conditions that the authors report. The same point is perhaps more relevant to table 3 - “children….. identified with intellectual disabilities in the HMDS but not identified in the IDEA database” - how many of these children were actually identified in HMDS as having intellectual disabilities as opposed to being identified as possibly having intellectual disabilities from the list of conditions? Additionally, related to this point, there is no information provided/referenced in the paper as to how complete/incomplete the IDEA database is with regards to cause of intellectual disabilities - it would be helpful if the authors commented on this. Please add to the methods that the relevant privacy approvals to use the data were in place.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Professor Roy McConkey
Institution and Country: Ulster University, Northern Ireland, UK
Please state any competing interests: None delcared

Please leave your comments for the authors below

Comment: This paper is predicated on a curious assumption that hospital morbidity data would provide a good indicator of the prevalence of ID in a childhood population (Page 4 lines 42 to 49). The authors cite no past research to support this proposition although the studies they do reference, indicate that only a minority of children known to have ID will have been admitted to hospital (albeit at a higher rate than their non-ID peers). Indeed their own study bears this out as shown in Figure 1; with around 18% of children with ID having a recorded hospital admission in a 27 year period.

Response: We apologise if we had not previously made the point of this paper as clear as it should be. We did not intend to suggest that hospital morbidity data be used as a single data source for studies investigating the prevalence of ID. As this was not our proposition we did not cite any such research but instead a meta-analysis illustrating that many different data sources have been used for this purpose. In fact our current study shows that of all children born in the period 1983-2010 and identified with ID in the IDEA database (n=10593), 92% (n=9740) had had a hospital admission not 18% as interpreted by reviewer one. What Figure 1 actually shows is that of the 9740 children with ID (as identified from the ID Database) admitted to hospital, only 1435 (14%) had an ICD code in the hospital morbidity database which could identify them as having ID. Therefore in the absence of a database recording intellectual disability in the population through linkage to both disability services and education, such as the IDEA database, the number of children identified through hospital morbidity coding alone could be a gross underestimate.
We have now made it clear in the results that 92% of individuals with ID in the IDEA Database have had a hospital admission in the study period and that this is the group that has been linked to the hospital morbidity database records. We hope this will have clarified any misunderstanding.
The aim of this paper was not, as we explained above, about estimating prevalence but simply to investigate how well hospital morbidity data identified ID in its coding, as stated in the Abstract Objective: “To investigate how well intellectual disability can be ascertained using hospital morbidity data compared with a population-based data source.” We have therefore removed the wording in the Introduction around use of hospital data as well as other administrative datasets in prevalence studies as this was not the focus of the paper and was obviously confusing.

We realise we should have made this clearer and have made appropriate changes to the wording in both the Introduction and Results.

Comment: Moreover the small number of children aged over 12 months identified solely in hospital admissions (n=354) would make little difference, I suspect, to the prevalence rates of ID within Western Australia. This is information that the authors do not include although it gets a fleeting mention in the discussion. Thus the extensive (and presuming time-consuming) data analysis only serves to confirm an unlikely proposition (acknowledged by the authors on page 8; lines 55 – 57 and again on page 12, lines 43-48).

Response: As already explained, we were not interested in the prevalence rates of ID using hospital morbidity data as we do have ascertainment through the ID Database.[1] What we wanted to acknowledge, however, was that there was a proportion of children "apparently" with an ID who were only identified in hospital morbidity data but as the reviewer points out this represented a very small number. What we did want to investigate were the characteristics of this small number of children in order to understand why they may be missing from the ID database. We found that one third had died before one year of age, and of the remaining 354 many were identified using specific morbidity diagnostic codes such as neurofibromatosis or tuberous sclerosis, which may not always be consistent with ID (Table 3). Our focus was more on the remaining children whom we knew had an ID from their existence in the IDEA database (n=8305), but may not have been allocated any ID code in hospital morbidity data.

Comment: The detailed comparisons that the authors present of the 'cause of disability' codings within the hospital sample and the larger ID population (Table 2) is of limited usefulness as no information is provided as to who undertook the codings in the two databases and the criteria used. Also it seems that only one 'cause code' was entered per case (based on totals given in Table 2) which can be a further source of inconsistency across the two systems. Indeed the authors go on to recognise in the discussion two further sources of error in the hospital codings: some children with ID will have been missed and children with particular syndromes may be assumed to have ID when this may not have been the case. Likewise the IDEA system misses children with ID who do not use/require services; especially younger children undergoing assessment.

Response: We apologise that it was not clear that the cause of intellectual disability was determined only from information in the IDEA database, essentially through medical records retained by the Disability Services Commission.
The purpose of the analysis presented in Table 2 was to examine the children in the ID Database who survived past one year and whether or not they had also been identified with ID in the hospital morbidity ICD codes. This allowed us to determine the likelihood of a child being allocated a hospital morbidity code for ID, based on the aetiology of ID.
The coding was carried out by medical doctors with long-standing clinical experience in intellectual disability as well as a clinician researcher with an extensive portfolio of research into intellectual disability.[2] Rightly so we believe, we have acknowledged any weaknesses and potential sources of error in our study as well as its strengths.
One such error that was acknowledged was that specific hospital morbidity codes for ID include some syndromes, e.g. Prader Willi syndrome, where ID is not present in 100% cases. However the number of such cases misidentified as ID would be small. The interesting point we found however was that only 14% of children known to have an ID were actually allocated any ID code in the hospital morbidity system.

We agree there is also the possibility that some children correctly coded with intellectual disability in hospital data are missing from IDEA, but have demonstrated this to be a small proportion of children. There is obviously the possibility that the IDEA database will not ascertain every child with ID, however our linkage to Education sources provides almost a third of the cohort and extends to those not requiring or accessing disability services.
We have now expanded our Methods section to provide more information about how the cause of disability was determined.
We have also changed the title of Table 2 to make this clearer.
We have added further discussion on the number of individuals identified through the HMDS codes but were not in IDEA, as this provides useful insight into the completeness of the IDEA database.

Comment: In sum, this study seems misdirected and not a good use of the human and financial resources incurred in data linkage studies. Rather a better focus would have been to explore the number and nature of hospital admissions of children with ID and possible changes over the years. Also the implications for primary care services that might have prevented or reduced the need for admissions perhaps could be identified as well as the implications for improved care in hospital.

Response: We apologise that we have not made the point of this study clearer- in that we wished to simply investigate how well hospital morbidity data could be used to identify ID by researchers who may only have this resource available. Our linked dataset has already been used for multiple studies, for example, investigation of prevalence,[1, 3] survival,[4, 5] maltreatment risk,[6] and the patterns of hospital admissions of children with ID[7, 8] while further research is currently underway investigating impact of certain interventions on hospitalisations.

Reviewer: 2
Reviewer Name: Sally-Ann Cooper
Institution and Country: University of Glasgow, UK
Please state any competing interests: None declared

Please leave your comments for the authors below

The study is clearly written and interesting. There are two main issues that the authors should address prior to publication.

1. It is obvious that hospital admission data cannot provide a credible population prevalence for intellectual disabilities. There are several reasons for this, including, a. only some infants/children require hospital admission, b. those that do are likely to have the reason for their admission recorded rather than whether they have intellectual disabilities (which one would expect to be recorded more hapazardly if at all), and c. the infant's/child's intellectual disabilities may not have been identified by the time of admission as some are not identified until school age. So what the study achieves is a quantification of the extent to which hospital admission is an inadequate source for identification of population prevalence of intellectual disabilities - they found only 14% of children thought to have intellectual disabilities could be identified in this way. I think the study would be more credible if the authors are clearer about this. In particular, paragraph 2 of the introduction needs rewriting. The first half of the sentence "on account of high likelihood of hospitalisations health data have often been used as a source of ascertainment for prevalence of intellectual disabilities, as have other administrative datasets relating to education or service provision as well as household sampling.." has not been justified by the references cited, and I suspect is actually incorrect.
That is, the first part of this sentence on health data has not been justified/is not correct (contrary to the statement regarding administrative education and other service datasets and household sampling

which is quite different and is justified by the meta-analysis quoted, and indeed the “gold standard” used in this study (IDEA) relates to administrative education and service data). The authors should rewrite this paragraph for greater accuracy so that their study has more credibility overall.

Response: We thank you for your comments and are pleased that you have understood that the paper demonstrates, as you have stated. “a quantification of the extent to which hospital admission is an inadequate source for identification of intellectual disabilities.” We hope that we have already made clear through our response to reviewer one that our aim was “To investigate how well “or how poorly” intellectual disability can be identified using hospital morbidity data.”
As also explained in our response to reviewer one, we have now included information on the proportion of children with ID (as identified from the ID database) who were hospitalized one or more times during the study period (92%).
We have rewritten paragraph 2 of the introduction and made changes to remove the wording around use of hospital data as well as other administrative datasets in prevalence studies as this was not the focus of the paper and we apologise for the unintended implication that health data alone would be used for prevalence estimates.

2. As the authors comment on in the discussion, the codes they use in HMDS for identification of children who might have intellectual disabilities are over-inclusive as not all children with these conditions will have intellectual disabilities e.g. marfan’s syndrome. They should therefore also be more precise in their descriptions/reporting in tables 2 and 3 and the text relating to these tables. Table 2 “children in the IDEA database……..identified/not identified with intellectual disabilities in HMDS…..” does not provide data on those identified/not identified with intellectual disabilities in HMDS, it provides data on those with the codes they have used as potential proxies for intellectual disabilities - unless I have misunderstood the methods/the methods have not been clearly written. Please clarify if these children have been identified/not identified with intellectual disabilities or if this refers to the longer list of conditions that the authors report.

Response:
Table 2 includes all children identified with ID through the IDEA database and surviving one year (n=9704) grouped as to whether or not they were also identified with ID from the HMDS ICD codes. We acknowledge that it was not clear that the cause of ID in Table 2 has been determined ONLY from the medical information in the IDEA database, not from the hospital morbidity codes (“potential proxies for ID”), where available. We have made changes to the Methods to clarify how cause of ID was determined and modified the Table 2 title. As noted in the following response further information on cause of ID (from IDEA) for those also identified through HMDS has also been added.

The same point is perhaps more relevant to table 3 - “children….. identified with intellectual disabilities in the HMDS but not identified in the IDEA database” - how many of these children were actually identified in HMDS as having intellectual disabilities as opposed to being identified as possibly having intellectual disabilities from the list of conditions?

Table 3 refers only to cases identified with ID using the HMDS codes and who were not identified by IDEA (n=478). From Figure 1 we can see that this represents 25% (478/(478 + 1435) of all children identified through the ICD codes (“potential proxies of ID”) and the remaining 75% (n=1435) had their ID confirmed through IDEA. However we have not specifically compared the cause of ID from IDEA with the ICD code identifying these 1435 children in the HMDS data. Those who survived past one year of age (n=1412) are included in Table 2 and we have shown that these children were most likely to have a cause of ID (as determined from IDEA medical information) of Down syndrome (42%, 589/1412). We have now also included this information in Results and modified the Table 3 title to clarify this.

Additionally, related to this point, there is no information provided/referenced in the paper as to how complete/incomplete the IDEA database is with regards to cause of intellectual disabilities - it would be helpful if the authors commented on this.

We agree there is some important information relevant to the completeness of IDEA ascertainment which should be reported. If we assume that all cases in Table 3 did have ID (n=478), then IDEA would represent 95.7% of true prevalence. With the assumption that those who died under one year would not be able to be ascertained (n=124, of whom the majority died under one month) then IDEA would represent 96.8%.
We have added this information to Results and Discussion.
The medical information on cause of ID in the IDEA database is limited to those referred to Disability Services Commission (DSC), so for about a third of the children ascertained only from education sources, as well as a small proportion of children with no medical information from DSC, cause of ID is described as “Unassessed”.
This has also been added to Methods.

Comment: Please add to the methods that the relevant privacy approvals to use the data were in place.

Response: This has been added to the Methods.

References
1. Bourke J, de Klerk N, Smith T, Leonard H. Population-Based Prevalence of Intellectual Disability and Autism Spectrum Disorders in Western Australia: A Comparison With Previous Estimates. Medicine (Baltimore). 2016;95(21):e3737.
2. Leonard H, Wen X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. Ment Retard Dev Disabil Res Rev. 2002;8(3):117-34.
3. Leonard H, Petterson B, Bower C, Sanders R. Prevalence of intellectual disability in Western Australia. Paediatr Perinat Epidemiol. 2003;17(1):58-67.
4. Glasson EJ, Jacques A, Wong K, Bourke J, Leonard H. Improved Survival in Down Syndrome over the Last 60 Years and the Impact of Perinatal Factors in Recent Decades. J Pediatr. 2016;169:214-20 e1.
5. Bourke J, Nembhard W, Wong K, Leonard H. Twenty five year survival in children with intellectual disability in Western Australia. J Pediatr. Forthcoming 2017.
6. Maclean MJ, Sims S, Bower C, Leonard H, Stanley FJ, O'Donnell M. Maltreatment Risk Among Children With Disabilities. Pediatrics. 2017;139(4).
7. Bebbington A, Glasson E, Bourke J, de Klerk N, Leonard H. Hospitalisation rates for children with intellectual disability or autism born in Western Australia 1983-1999: a population-based cohort study. BMJ Open. 2013;3(2).
8. Fitzgerald P, Leonard H, Pikora TJ, Bourke J, Hammond G. Hospital admissions in children with Down syndrome: experience of a population-based cohort followed from birth. Plos One. 2013;8(8):e70401.

VERSION 2 – REVIEW

REVIEWER	Roy McConkey Ulster University, Northern Ireland.
REVIEW RETURNED	30-Oct-2017

GENERAL COMMENTS	I appreciate the authors' response to my previous comments and note that they have clarified the aim of the paper to focus on the coding of ID within hospital datasets, rather than the use of hospital data in estimating the prevalence of ID. This brings into particular focus my previous comment regarding who undertook the coding of ID (and other ICD codes) in the hospital dataset and also within the IDEA data. The paper is still silent on this crucial matter. I would have expected at least a description of the coding processes but better still would be some evidence of cross-checking to confirm the assigned codes. Indeed I suspect that a more qualitative study with hospital personnel would be required to understand more fully their reasons for coding or not coding a child as having ID, and to identify if coding with ID would indeed improve the hospital care as the authors' suggest (p.15 lines 53 to 57). The authors have clarified that a high proportion of children aged over one year on the IDEA database were hospitalised during the study period (n=9,740). However only 19% were recorded on the hospital dataset as having ID. The authors then compare the causes of ID coded for children identified as having ID in both datasets compared to the causes recorded for children only on IDEA – presumably those who had a hospital admission? How was the cause of their ID established as shown in the second column of Table 2? To what extent did the hospital record for their admission provide this information (or could it?), or was it recorded in the IDEA dataset and if so who assigned these codes (which raises similar issues around cross-checking). In any case Table 2 is overly detailed and risks comparing percentages calculated on diverse but mostly small numbers. The authors rightly point out other difficulties about hospital-only codes imputing ID to certain conditions when in fact this may apply to only a proportion of persons so affected. This may become even more of an issue if greater emphasis is placed, as the authors suggest, on "Orphanet coding into ICD-11 will allow many more genetic ID syndromes to be specifically identified in hospital morbidity data" (p.14). ID is more a social disability as per the International Classification of Functioning (WHO) than a clinical condition. The discussion on page 15 harks back to the previous version of the paper that focussed on prevalence and has less relevance to the present redrafted paper. The authors' conclusions regarding hospital codings are however worthy of note – particularly hospital "coding practices which identify ID need to be better implemented". Hence my focus on auditing the current coding practices within datasets and reviewing their usefulness. 'Big data' studies cannot take for granted that the information they link is reliable and valid. I still remain unconvinced that a full paper is needed to question the value of hospital-datasets regarding ID when a letter to the editor might be more impactful.

REVIEWER	Sally-Ann Cooper University of Glasgow UK
REVIEW RETURNED	10-Oct-2017

GENERAL COMMENTS	The paper is an interesting addition to the existing body of literature, and has been improved through the revision.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Roy McConkey
Institution and Country: Ulster University, Northern Ireland.
Please state any competing interests: None declared

Please leave your comments for the authors below

Comment: I appreciate the authors' response to my previous comments and note that they have clarified the aim of the paper to focus on the coding of ID within hospital datasets, rather than the use of hospital data in estimating the prevalence of ID. This brings into particular focus my previous comment regarding who undertook the coding of ID (and other ICD codes) in the hospital dataset and also within the IDEA data. The paper is still silent on this crucial matter. I would have expected at least a description of the coding processes but better still would be some evidence of cross-checking to confirm the assigned codes.

Response: Within hospital settings in Western Australia (WA), as occurs internationally, clinical coding of all hospital separations, discharges and deaths is undertaken using the International Classification of Diseases (ICD) system. In WA this process is described as "translation of written clinical documentation about patient care into code format. A clinical coder is responsible for abstracting relevant information from the medical record and deciding which diagnoses and procedures meet criteria for coding as per Australian and WA Coding Standards. The coder then assigns codes for these diagnoses and procedures based on ICD-10-AM conventions and standards." [1] It is these ICD codes, which include specific codes for mental retardation as well as some conditions consistent with intellectual disability, (such as Down syndrome as listed in our Methods) that we and others [2,3] have used to identify intellectual disability in hospital morbidity data.
The IDEA dataset ascertains individuals meeting criteria for ID either from information through the Disability Services Commission (DSC) which also provides information on medical causes of ID where available; and the Department of Education (which does not provide any medical information). Therefore everyone in the IDEA dataset has a confirmed ID. At ascertainment medical information available through the DSC has been coded, over the 23 years included in this study, by the medical team at DSC according to an aetiological coding system based on that developed by AAMR,[4,5] which assign a cause of ID or according to conditions associated with ID. These codes have been consistently applied over the past 40 years to individuals in IDEA. We are confident about the use of these codes as they have been used in analyses for multiple peer reviewed studies, a sample of which is referenced here.[6-12]

Whilst these two coding systems, ICD and AAMR [5], have some common codes, (such as some of the conditions listed in Methods – Down syndrome, Prader Willi, Neurofibromatosis) the AAMR classification [5] has more detail than the available ICD codes and so cross checking of the codes assigned for those identified in both systems is not useful. We can see however that for those individuals with conditions such as Down syndrome, Prader Willi and Neurofibromatosis they are almost always identified as such in both systems. In looking at those with the ICD general codes for intellectual disability (ICD-9-CM 317-319; ICD-10-AM F70-F79) , we found, as expected, that there was a wide spread of AAMR codes designated to these children including autism, associated with epilepsy, familial ID, other syndromes, postnatal injury and many others.
In the Results section we have now included some text as follows, on the alignment of cause of ID, where it is feasible to do so,
"For the children who were identified through both IDEA and HMDS and survived one year of age (n=1412), n=623 had an ICD code for "mental retardation". The consensus of diagnosis between IDEA and the ICD codes for particular disorders was 80-98% for Down syndrome, Trisomy 18/13, Trisomy 9/8, Chromosomal deletions, Fragile X syndrome, Tuberous sclerosis and Prader-Willi syndrome; and less for Neurofibromatosis (63%) and Marfan syndrome (12.5%)."
We have also added some further information on coding practices for both hospital morbidity and the IDEA database in the Methods to enable a better understanding of the dataset we used.

Comment: Indeed I suspect that a more qualitative study with hospital personnel would be required to understand more fully their reasons for coding or not coding a child as having ID, and to identify if coding with ID would indeed improve the hospital care as the authors' suggest (p.15 lines 53 to 57).

Response: A qualitative study, although desirable, would not be feasible in most health systems given the rarity of ID and the multiple personnel involved in coding. However we agree it would be helpful to acquire greater insight into coding practices within the hospital system. The current study was undertaken to investigate how ID is currently recognised as a comorbidity in hospital separation data, given the increased likelihood of hospitalisation experienced by people with ID.[10]

Comment: The authors have clarified that a high proportion of children aged over one year on the IDEA database were hospitalised during the study period (n=9,740). However only 19% were recorded on the hospital dataset as having ID. The authors then compare the causes of ID coded for children identified as having ID in both datasets compared to the causes recorded for children only on IDEA – presumably those who had a hospital admission?

Response: Yes, the study only included children who had been hospitalised, which represented 92% of all children in the IDEA database. This is the second sentence of the Results. We then compared which of these children had also been identified with ID according to the hospital morbidity codes and looked at their characteristics.

Comment: How was the cause of their ID established as shown in the second column of Table 2?

Response: The process of defining a cause of ID for the IDEA Database has been undertaken over time by the medical team at the Disability Services Commission, often in consensus, based on the medical records of the individual person, using the AAMR codes. [5] For this analysis, given the number of individual codes, a regrouping of cause of ID into a smaller number of categories was undertaken based on Yeargin-Allsopp's groupings of biomedical or other causes [13], as listed in Table 2.

Comment: To what extent did the hospital record for their admission provide this information (or could it?

Response: Information from hospital records was not used for cause of ID in Table 2 as it was only available for a small proportion, however information on the alignment of diagnosis from hospital records with the causes from IDEA has been discussed above and text has been added to the manuscript.
, or was it recorded in the IDEA dataset and if so who assigned these codes (which raises similar issues around cross-checking)

Response: See above

Comment: In any case Table 2 is overly detailed and risks comparing percentages calculated on diverse but mostly small numbers.

Response: The information in Table 2 is intended as descriptive information only and statistical tests are not appropriate.

Comment: The authors rightly point out other difficulties about hospital-only codes imputing ID to certain conditions when in fact this may apply to only a proportion of persons so affected. This may become even more of an issue if greater emphasis is placed, as the authors suggest, on “Orphanet coding into ICD-11 will allow many more genetic ID syndromes to be specifically identified in hospital morbidity data” (p.14). ID is more a social disability as per the International Classification of Functioning (WHO) than a clinical condition.

Response: ID may be a social disability but its diagnosis is based on cognitive ability and adaptive behaviours [14] and the ability to recognise genetic syndromes into the future will allow for greater research into causes.

Comment: The discussion on page 15 harks back to the previous version of the paper that focussed on prevalence and has less relevance to the present redrafted paper.

Response: We have removed the discussion around other studies which have used ICD coding in health datasets as a method of estimating prevalence of ID.

Comment: The authors’ conclusions regarding hospital codings are however worthy of note – particularly hospital “coding practices which identify ID need to be better implemented”. Hence my focus on auditing the current coding practices within datasets and reviewing their usefulness. ‘Big data’ studies cannot take for granted that the information they link is reliable and valid.

ResponsE: I think this is the point we are making, that having the population-based IDEA database which ascertains individuals with a confirmed ID, has enabled us to identify and quantify the limitations in the “big data” hospital dataset as far as identifying people with ID. This does not undermine however the value to be gained from linking to hospital or birth data in order to investigate other determinants or outcomes for this population.

Comment: I still remain unconvinced that a full paper is needed to question the value of hospital-datasets regarding ID when a letter to the editor might be more impactful.

Response: We believe that information presented in this paper will be seen as valuable and “an interesting addition to the existing body of literature “ as suggested by the other reviewer.

Reviewer: 2
Reviewer Name: Sally-Ann Cooper
Institution and Country: University of Glasgow, UK
Please state any competing interests: None

Please leave your comments for the authors below
The paper is an interesting addition to the existing body of literature, and has been improved through the revision.

References
1 http://ww2.health.wa.gov.au/Articles/A_E/Clinical-Coding-Authority
2 Westerinen H, Kaski M, Virta L, Almqvist F, Iivanainen M. Prevalence of intellectual disability: a comprehensive study based on national registers. Journal of Intellectual Disability Research. 2007;51:715-25.
3 Lin E, Balogh R, Cobigo V, Ouellette-Kuntz H, Wilton AS, Lunsky Y. Using administrative health data to identify individuals with intellectual and developmental disabilities: a comparison of algorithms. J Intellect Disabil Res. 2013;57:462-77.
4 Luckasson, R., Borthwick-Duffy, S., Buntinx, W. H. E., Coulter, D. L., Craig, E. M., Reeve, A. I., et al. (2002). Mental retardation: definition, classification, and systems of supports. Washington, DC: American Association on Mental Retardation.
5 Heber, R. (Ed.) (1961). A manual on terminology and classification on mental retardation (rev. ed.). Washington, DC: American Association on Mental Deficiency.
6 Leonard H, Petterson B, De Klerk N, Zubrick SR, Glasson E, Sanders R, et al. Association of sociodemographic characteristics of children with intellectual disability in Western Australia. Soc Sci Med. 2005;60(7):1499-513.
7 Williams K, Leonard H, Tursan d'Espaignet E, Colvin L, Slack-Smith L, Stanley F. Hospitalisations from birth to 5 years in a population cohort of Western Australian children with intellectual disability. Arch Dis Child 90(12):1243-8. 2005;90(12):1243-8.
8 Leonard H, Glasson E, Nassar N, Whitehouse A, Bebbington A, Bourke J, et al. Autism and intellectual disability are differentially related to sociodemographic background at birth. PLoS ONE. 2011;6(3):e17875.
9 O'Leary C, Leonard H, Bourke J, D'Antoine H, Bower C. Intellectual disability: population-based estimates of the proportion attributable to heavy prenatal alcohol exposure. 2012, Developmental Medicine and Child Neurology, 2013. 55(3):271-7
10 Bebbington A, Glasson E, Bourke J, de Klerk N, Leonard H. Hospitalisation rates for children with intellectual disability or autism born in Western Australia 1983–1999: a population-based cohort study. BMJOpen, 2013. 3(2)
11 Maclean MJ, Sims S, Bower C, Leonard H, Stanley FJ, O'Donnell M. Maltreatment risk among children with disabilities. Pediatrics. 2017;139(4).
12 Bourke J, Nembhard W, Wong K, Leonard H. Twenty-Five Year Survival of Children with Intellectual Disability in Western Australia. Journal of Pediatrics. 2017 in press
13 Yeargin-Allsopp M, Murphy CC, Cordrro JF, Decoufle P, Hollowell JG. Reported biomedical causes and associated medical conditions for mental retardation among 10-year-old children, metropolitan Atlanta, 1985 to 1987. Dev Med Child Neurol. 1997;39:142-9.
14 Schalock R, Borthwick-Duffy S, Bradley V, Buntinx W, Coulter D, Craig E, et al. Intellectual disability: Definition, classification, and systems of supports. 11th ed. Washington, DC : American Association on Intellectual and Developmental Disabilities; 2010.